(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平10－257436

(43)公開日　平成10年(1998)９月25日

(51)Int.Cl.[6]		識別記号	FI		
H04N	5/93		H04N	5/93	Z
	5/92			5/92	H
	7/32			7/137	Z

審査請求　未請求　請求項の数７　ＯＬ　（全 14 頁）

(21)出願番号　特願平9－55340

(22)出願日　平成９年(1997)３月10日

(71)出願人　391023987
松下　温
東京都新宿区喜久井町36

(71)出願人　392008231
岡田　謙一
東京都文京区本郷４－25－12

(72)発明者　松下　温
東京都新宿区喜久井町36

(72)発明者　岡田　謙一
東京都文京区本郷４－25－12

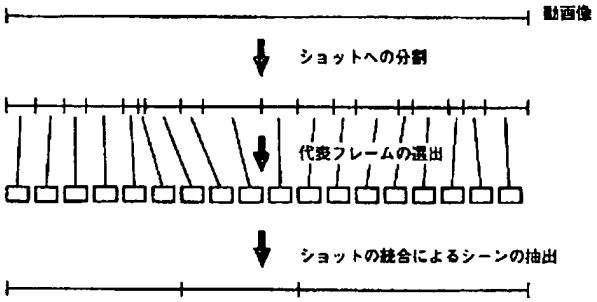
(74)代理人　弁理士　鈴木　正次

(54)【発明の名称】　動画像の自動階層構造化方法及びこれを用いたブラウジング方法

(57)【要約】

【課題】　この発明は、符号化した動画像を自動階層構造化し、この動画像とその解析データを基にしてビデオブラウザを得ることを目的としたものである。

【解決手段】　動画像を符号化し、該符号化された動画像を各ショットに分割し、ついで分割されたショット毎の類似度を用い、ショットを統合してシーンを抽出処理することを特徴とした動画像の自動階層構造化方法。動画像を符号化し、該符号化された動画像を各ショットに分割し、ついで分割されたショット毎の類似度を用い、ショットを統合してシーンを抽出処理して動画像を自動階層構造化し、この階層構造化されたデータを用いて動画像全体の内容把握、所望のシーンまたはショットの検出を容易にすることを特徴とした動画像のブラウジング方法。



【特許請求の範囲】
【請求項１】　符号化された動画像を各ショットに分割し、ついで分割されたショット毎の類似度を用い、ショットを統合してシーンを抽出処理することを特徴とした動画像の自動階層構造化方法。
【請求項２】　動画像の符号化は、ＭＰＥＧによるものとすることを特徴とした請求項１記載の動画像の自動階層構造化方法。
【請求項３】　符号化された動画像からショットを検出する際に、ＭＰＥＧの特徴を利用して高速に処理することを特徴とした請求項１記載の動画像の自動階層構造化方法。
【請求項４】　ショット同士の類似度の算出に際し、代表フレームを抽出することを特徴とした請求項１記載の動画像の自動階層構造化方法。
【請求項５】　ショット間の類似度をファジィ推論により求めることを特徴とした請求項１記載の動画像の自動階層構造化方法。
【請求項６】　シーンの抽出処理は、定義されたショット間の結合度により求めることを特徴とした請求項１記載の動画像の自動階層構造化方法。
【請求項７】　符号化された動画像を各ショットに分割し、ついで分割されたショット毎の類似度を用い、ショットを統合してシーンを抽出処理して動画像を自動階層構造化し、この階層構造化されたデータを用いて動画像全体の内容把握、所望のシーンまたはショットの検出を容易にすることを特徴とした動画像のブラウジング方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】この発明は、符号化した動画像を自動階層構造化し、この動画像とその解析データを基にしてビデオブラウザを得ることを目的とした動画像の自動階層構造化方法及びこれを用いたブラウジング方法に関する。
【０００２】
【従来の技術】現在動画像情報は、ビデオの単純再生の域を脱していない。即ち動画像はフレーム単位でとらえられており、撮影時に符号を付した場合には、その符号によって抽出することができると共に、再生時間をファクターとして所望のフレームを検出又は再生するなど特別な関係が明らかな場合に限り、当該フレームを抽出することができている。
【０００３】
【発明により解決すべき課題】然し乍ら何等の符号を付することなく、与えられる動画像情報から所望のフレームを抽出することは極めて困難であり、時間的制約があれば、抽出不可能となる。例えば一般に１フレームは３０分の１秒であるから、１分間に１８００フレーム、１時間に１０８０００フレームとなる。
【０００４】そこで前記従来の一般動画像情報から任意のフレームを短時間で抽出することができない問題点があった。
【０００５】
【課題を解決する為の手段】然るにこの発明は、符号化された動画像を各ショットに分割し、ショット毎の類似度を用いてショットを統合し、シーンを抽出することにより自動階層構造化し、このデータを用いて動画像のブラウジングツールを作成することによって前記従来の問題点を解決したのである。
【０００６】即ちこの発明は、符号化された動画像を各ショットに分割し、ついで分割されたショット毎の類似度を用い、ショットを統合してシーンを抽出処理することを特徴とした動画像の自動階層構造化方法であり、動画像の符号化は、ＭＰＥＧによるものとすることを特徴としたものである。また符号化された動画像からショットを検出する際に、ＭＰＥＧの特徴を利用して高速に処理することを特徴としたものであり、ショット同士の類似度の算出に際し、代表フレームを抽出することを特徴としたものである。次にショット間の類似度をファジィ推論により求めることを特徴としたものであり、シーンの抽出処理は、定義されたショット間の結合度により求めることを特徴としたものである。更に他の発明は、符号化された動画像を各ショットに分割し、ついで分割されたショット毎の類似度を用い、ショットを統合してシーンを抽出処理して動画像を自動階層構造化し、この階層構造化されたデータを用いて動画像全体の内容把握、所望のシーンまたはショットの検出を容易にすることを特徴とした動画像のブラウジング方法である。
【０００７】前記における符号化はＭＰＥＧ１の圧縮アルゴリズムによる。ここにＭＰＥＧ１の正式名称は「Ｃｏｄｉｎｇ　ｏｆ　ｍｏｖｉｎｇ　ｐｉｃｔｕｒｅｓ　ａｎｄ　ａｓｓｏｃｉａｔｅｄ　ａｕｄｉｏ　ｆｏｒ　ｄｉｇｉｔａｌ　ｓｔｏｒａｇｅ　ｍｅｄｉａ　ａｔ　ｕｐ　ｔｏ　ａｂｏｕｔ　１．５　Ｍｂｉｔ／ｓ」である。
【０００８】前記ハイブリッド符号化は、ＤＣＴと量子化、動き補償フレーム間予測及びエントロピー符号化により行うが、前記個々の方法は公知の方法であるから詳細な説明は省略する。
【０００９】次にＭＰＥＧ１符号化、復号システムを図１について説明する。ビデオ入力は、前処理を経てビデオ符号化器に入り、ついでシステム多重化を経て、蓄積メディアに入りついでシステム多重分離した後、ビデオ復号器に入り前処理してビデオ出力となる。
【００１０】前記ＭＰＥＧ１ビデオ符号化器は図２の入力画像がデータに処理される。またＭＰＥＧ１ビデオ復号器は、図３のように入力バッファが表示バッファに処理される。
【００１１】前述したように、ＭＰＥＧ１はＣＤ－ＲＯ

Mなどの蓄積メディアに用いることが目的である。蓄積メディアでは、早送り、巻戻し、途中からの再生、逆転再生などのトリックモードが必要とされる。このようなトリックモードを実現するため、ＭＰＥＧ１ではグループオブピクチャー（Ｇｒｏｕｐ　ｏｆ　Ｐｉｃｔｕｒｅｓ、以下ＧＯＰという）構造が取られている。

【００１２】ＭＰＥＧ１では、符号化された画像データは、前後の画面データをもとにして作られているために、１画面だけで完結した情報にはならない。このために、何枚かの画面データをひとまとまりにしたＧＯＰを単位として、ランダムアクセスを可能にしている。つまり、ＧＯＰの中に少なくとも１枚は、前後画面の情報を利用せず１枚だけで閉じた画面データ（Ｉピクチャ）を必ず含むようにすることで、このデータを元にＧＯＰ内の他の画面データの再生が可能となる。なお、１つのＧＯＰは、通常１５枚程度のピクチャをグループ化することが多い（図４）。

【００１３】ＭＰＥＧ１では、過去再生画像からの順方向予測と未来再生画像からの逆方向予測の両方を行っている。これを双方向予測という。

【００１４】双方向予測を実現するため、ＭＰＥＧ１では、Ｉピクチャ、Ｐピクチャ、Ｂピクチャの３つのタイプの画像を規定している。

【００１５】これらの他に、Ｄピクチャ（ＤＣ符号化画像）が規定されている。これは、フレーム内の情報のみで符号化され、ＤＣＴ係数の内のＤＣ成分のみで構成されており、他の３種類のピクチャタイプと同じシーケンスに共存することはない。

【００１６】ＭＰＥＧ１では、双方向予測を行うＢピクチャが導入されることによって、予測効率が大きく向上し、高圧縮時の画質向上に役立っている。

【００１７】画像データは、図６に示すように、シーケンス、ＧＯＰ、ピクチャ、スライス、マクロブロック（ＭＢ）、ブロックの６層の階層構造から成っている。

【００１８】前記シーケンス層とは、一続きの映像を表現するビットストリームは、シーケンスヘッダで始まり、その後に１個または数個のＧＯＰが続き、最後に１個のシーケンスエンドコードで終了する。どのＧＯＰの直前にもシーケンスヘッダを置くことができるが、一続きの映像中のシーケンスヘッダでは量子化マトリクス以外のデータ要素は全て最初のシーケンスヘッダと同じである必要がある。

【００１９】これによってシーケンス途中へのランダムアクセスが可能になる。

【００２０】またＧＯＰ層とはＧＯＰをひとつ含む。

【００２１】次にピクチャ層とはＩピクチャ、Ｐピクチャ、Ｂピクチャ、Ｄピクチャのいずれかを１枚含む。

【００２２】またスライス層とは、スライスは、画像の左上から始まってラスタスキャン順に右下に続く一連の任意個のマクロブロックの集まりである。スライス間には重なりやすき間を持たせることはできないが、スライスの位置は画面ごとで異なってもよい。スライスのデータの先頭には同期信号が割り当てられるため、復号時にデータの読みだし誤差があっても次のスライスで同期を回復できる利点がある。またスライスのデータの復号はそのスライスだけ独立して行えるため、復号の高速化のためにスライス単位に並列処理が可能である。

【００２３】次にマクロブロック層とは、マクロブロックは１６画素×１６ラインの輝度成分と、画像中で空間位置が対応する８画素×８ラインの２つの色差成分で構成されている。ひとつのマクロブロックは４個の輝度ブロックと２個の色差ブロックからなる。マクロブロック中でのブロック順序と配置は図６のとおりである。このマクロブロックを単位に動き補償およびフレーム間予測は行われる。

【００２４】更にブロック層とは、８画素×８ラインからなる輝度成分または色差成分で構成されるＤＣＴ処理単位である。

【００２５】

【発明の実施の形態】この発明は、符号化された動画像を各ショットに分割し、ショット毎の類似度を用いてショットを統合し、シーンを抽出するようにした動画像の自動階層構造化の方法である。

【００２６】また前記階層構造化されたデータを用いて動画像の全体の内容を把握し、また所望のシーン、ショット又はフレームの検索を容易にした動画像のブラウジング方法である。

【００２７】前記この発明により、動画像の内容を把握したり、所望の場面を検索することが極めて容易となった。

【００２８】

【実施例】この発明の実施例を図面に基づいて説明する。

【００２９】まず、物理的な特徴量によって抽出が可能であり、検出が比較的容易な、ショットへと動画像を分割する。そして、分割されたショット間の類似度によって、ショットを統合することでシーンの抽出を行う。この際、ショットという動画像のままでは扱いにくいため、ショット中から代表フレームをいくつか選び出す（図７）。

【００３０】一般に、動画像はデータ量が多いため、その処理量は膨大なものとなる。さらに、符号化された動画像の場合、復号化が必要となるため、さらに処理量は増大する。

【００３１】この発明では、ＭＰＥＧ１動画像を完全に復号化することなく、必要最小限の情報のみを復号化することで、高速な処理を可能としているので、フレーム間予測やＤＣ成分の復号化による簡略画像の取得といったＭＰＥＧ１の符号化アルゴリズムの特性を利用して処理量を軽減している。そこでＭＰＥＧ動画像中のＩピク

チャの簡略復号化とフレームの比較の方法について述べ、その後流れに沿って各処理の詳細を説明する。

【0032】動画像は多数の静止画像（フレーム）によって構成されている。したがって、動画像の解析には各フレームの画像情報は必要不可欠である。しかし、MPEG1動画像の復号化は、比較的処理量が多く、高速な処理を実現することは難しい。

【0033】そこで、すべてのフレームではなくIピクチャだけを復号化し、さらに、完全な復号化ではなく、簡易復号化によって原フレームの縮小画像を得る。この簡易復号化は、DCT係数のDC成分を復号化すると元のブロックの平均色が得られることを利用する。つまり、各ブロックのDCT係数のうちDC成分だけを復号化し、得られた平均色で各ブロックを代表させた画像を作るのである（図8）。

【0034】このようにして得られた画像をDC画像と呼ぶこととする。各ブロックの大きさは8×8であるから、DC画像は縦横それぞれ原画像の1／8の大きさとなる。

【0035】BピクチャおよびPピクチャの復号には、自分自身だけでなく、動きベクトル情報および参照先のピクチャなど直接使わない情報の復号が必要となるが、Iピクチャはフレーム内で閉じた符号化がなされているので、そうした情報の復号化の必要はない。また、Iピクチャ中のイントラ（Intra）マクロブロックのDC成分は、計算量の多いIDCTを行うことなく下式（1）によって復号化することができる。したがって、DC画像は非常に高速に得ることができる。

【0036】

【数1】

$$\begin{aligned} Y_k &= (DY_k)/8 \\ Cb_{k'} &= (DCb_{k'})/8 \\ Cr_{k'} &= (DCr_{k'})/8 \end{aligned} \qquad (1)$$

【0037】ここで、Y_k、$Cb_{k'}$、$Cr_{k'}$は各ブロック（$_k$、$_{k'}$はブロック番号）の平均色の輝度および色差成分、DY_k、$DCb_{k'}$、DCrは各ブロックのDC成分である。

【0038】実際に、約30分のMPEG1動画像（GOPは図5のタイプのもの）について、全てのフレームを復号した場合と、IピクチャのDC画像だけを復号した場合の処理時間を表1に示す。全てのフレームを復号する場合に比べ、約1／20の処理時間で復号できることがわかる。

【0039】

【表1】

表1　復号時間の比較

画像の長さ	1832.0
全てのフレームを復号	621.7
Iピクチャだけを復号	154.3
DC画像だけを復号	32.6

【0040】また、図9に、DC画像の例と、その原画像を示す。

【0041】フレーム間の比較に用いる類似度は、画素値の比較と色ヒストグラムの比較に大別される。色ヒストグラムによる比較はカメラや被写体の動きに影響を受けにくいために、類似度として用いるには都合が良いが、半面、空間的な情報を全く含まないために全く違う画像が同じ色ヒストグラムを持つ場合が問題となる。色ヒストグラムによる比較に空間的な情報を持たせようとする試みはいくつかなされているが、いずれもある程度複雑な処理を必要とする。

【0042】この発明では、色ヒストグラムによる距離として式（2）の$D_{histarea}$を、画素値による距離として式（3）のD_{pixsum}を用い、このふたつの値を組み合わせて類似度を算出することで処理の単純さを損なわずに空間的な情報を加味した類似度を求める。

【0043】

【数2】

$$D_{histarea} = 1 - \sum_{n=0}^{N_{bin}} \min(I_n, J_n) \Big/ \sum_{i}^{N_{bin}} I_n \qquad (2)$$

【0044】

【数3】

$$D_{pixsum} = \frac{1}{H \cdot W} \sum_{i=0}^{H} \sum_{j=0}^{W} |a_{ij} - b_{ij}| \qquad (3)$$

【0045】ここで、ふたつの値から類似度を算出する手法として、簡略化ファジィ推論を用いる。ファジィ推論を用いる事で、色ヒストグラムによる距離および画素値による距離と、画像の類似度の関係を厳密に定式化する事なく記述できる。また、簡略化ファジィ推論による推論は単純で、高速に実行できる。

【0046】このとき用いるファジィルールは以下式（4）の通りである。

【0047】

【数4】

$$\text{rule i:}\quad \text{IF}\quad D_{histarea}\ \text{is}\ A_a \quad D_{pixsum}\ \text{is}\ B_b \quad \text{THEN}\quad s_i = c_i \qquad (4)$$
$$a, b = \text{small, medium, large}$$
$$i = 1, 2, \cdots, 9$$

【００４８】ここで、ｉはルール番号、Ｉはルール数、c_i は後件部を表す実数値であり、［０、１］の値をとる。また、A_a 、B_b はそれぞれその特徴値のメンバシップ関数であり、各特徴値ごとに図１０のような“ｓｍａｌｌ”、“ｍｅｄｉｕｍ”、“ｌａｒｇｅ”の3つのメンバシップ関数を設定する。

【００４９】このルールに対する適合度を式（５）により求め、次に式（６）で最終的な推論結果、すなわち画像間の類似度ｓを求める。なお、ｓは［０、１］の値を取る。

【００５０】

【数５】

$$w_i = \mu_{A_a}(D_{histarea}) \cdot \mu_{B_b}(D_{pixsum}) \qquad (5)$$

【００５１】

【数６】

$$s = \sum_{i=1}^{I} w_i c_i \Big/ \sum_{i=1}^{I} w_i \qquad (6)$$

【００５２】ショットの検出はショットの間のカット点の検出を行う。

【００５３】カット点の検出とは、フレーム間の相関の低い点を検出する作業に他ならないが、この発明では、フレームの特徴量を直接比較して相関を調べるのではなく、ＭＰＥＧ１の符号化の様子から相関を調べ、カット点を検出する。つまり、ＭＰＥＧ１において、フレーム間の相関から予測によって圧縮が行われていることを利用し、逆に、予測の行われ方を調べることでフレーム間の相関を調べるのである。

【００５４】新たにフレームの特徴量を調べることなく、ＭＰＥＧ１の符号化情報を利用することで、計算量が少なくて済み、また、すべての情報を復号化する必要がないため、高速な処理が可能である。

【００５５】処理手順としては、まず、Ｂピクチャにおける参照の様子からカット点を検出し、さらにＰピクチャでの参照、Ｉピクチャの変化を調べて確認を行う。図１１のようなＮ＝１５、Ｍ＝３のＧＯＰを例に説明する。

【００５６】Ｂピクチャでは、前後両方のＩまたはＰピクチャから参照を行なっている。即ちＢピクチャ中には、一般的に（ＩＭＢ）、（ＦＭＢ）、（ＢＭＢ）、（ＢｉＭＢ）の４種類のマクロブロックＭＢが存在し、それぞれ、参照を全く行わないＩＭＢか、順方向ＦＭＢ、逆方向ＢＭＢ、双方向ＩｎＭＢの参照を行っている。このときの参照の様子は図１２のようになる。

【００５７】ショットの中、すなわちフレーム間の相関が高い場合には、過去および未来への参照の数はほぼ等しいが、参照するフレームとの間にカット点が存在すると、過去または未来へ依存が大きく偏り、マクロブロックの構成に偏りが生じる。このときの様子を図１２（ｂ）（ｃ）（ｄ）に示す。ただし、図１２は極端な場合であり、実際にはカット点を越えた参照が完全になくなる訳ではない。

【００５８】このことからわかるように、Ｂピクチャのマクロブロックタイプの構成から、Ｂピクチャの前後フレームへの参照の様子を判断することができる。これをＢピクチャの依存度ｒｅｌａｔとして式（７）のように定義する。

【００５９】

【数７】

$$relat = \frac{N_F - N_B}{N_{Bi}} \qquad (7)$$

【００６０】ただし、N_F 、N_B 、N_{Bi}はそれぞれＢピクチャに含まれるＦＭＢ、ＢＭＢ、ＢｉＭＢの数である。

【００６１】ｒｅｌａｔは、N_F とN_B の差が大きく、またN_{Bi}が少ない程、その絶対値が大きくなり、参照の偏りが大きいことを示す。

【００６２】さて、図１１のようなＧＯＰにおいて、ふたつのＰピクチャ（またはＩピクチャ）とそれに挟まれたふたつのＢピクチャに注目し（例えばf_7 、f_8 、f_9 、f_{10}）、これをP_1 B_2 B_3 P_4 と表すことにすると、全てのカット点は必ずP_1 ｜B_2 B_3 P_4 、P_1 B_2 ｜B_3 P_4 、P_1 B_2 B_3 ｜P_4 のいずれかの形で現れる（｜はカット点を表す）。これらはそれぞれ図１２の（ｂ）（ｃ）（ｄ）に対応する。このとき、どの場合でも、図１２からわかるとおり、B_1 、B_2 の両方に参照の偏りが生じ、依存度の絶対値が大きくなる。

【００６３】そこで、次式（８）を満たせば、P_1 ｜B_2 B_3 P_4 、P_1 B_2 ｜B_3 P_4 、P_1 B_2 B_3 ｜P_4

のいずれかの形でカット点が存在すると判断する。

【0064】

【数8】

$$|relat_{B_1} \cdot relat_{B_2}| > threshold_B \qquad (8)$$

【0065】次に、P_1 | B_2 B_3 P_4 、P_1 B_2 | B_3 P_4 、P_1 B_2 B_3 | P_4 のどのパターンかを判断し、正確なカット点を決定する。式7からもわかるとおり、ｒｅｌａｔは、過去からの参照が多いと正、未来からの参照が多いと負の値を取る。これを利用して、式（９）のようにカット点を決めることができる。

【0066】

【数9】

$$\begin{aligned} &\mathbf{P_1|B_2B_3P_4} \quad \text{if} \quad relat_{B_1} < 0, \quad relat_{B_1} < 0 \\ &\mathbf{P_1B_2|B_3P_4} \quad \text{if} \quad relat_{B_1} > 0, \quad relat_{B_1} < 0 \\ &\mathbf{P_1B_2B_3|P_4} \quad \text{if} \quad relat_{B_1} > 0, \quad relat_{B_1} > 0 \end{aligned} \qquad (9)$$

【0067】以上のようにして、Ｂピクチャの参照情報からカット点を検出することができる。

【0068】Ｂピクチャの参照による検出だけでは、ノイズや、カメラの前を物体が横切るなど瞬間的な画面の変動がある際に、誤検出が発生することがある。これは、Ｂピクチャと参照先のピクチャとの距離が短いためと考えられる。そこで、Ｂピクチャだけではなく、より遠いピクチャを参照するＰピクチャの参照情報を利用して結果の確認を行う。

【0069】Ｐピクチャは、参照しているＩまたはＰピクチャとの間にカット点が存在すると、参照がほとんどできなくなるため、ＩＭＢの数が増加するはずである。そこで、次式（１０）を満たす場合は、間にあるＢピクチャから求めたカット点は誤検出であるとみなし、これを除去する。

【0070】

【数10】

$$\frac{N_I}{N} < threshold_P \qquad (10)$$

【0071】ただし、N_I はＰピクチャに含まれるＩＭＢの数、Ｎはピクチャ中の全ＭＢの数である。

【0072】Ｐピクチャよりさらに離れたＩピクチャどうしの比較を使った確認を行う。

【0073】また、図１１のタイプのＧＯＰの場合、f_3 のＩピクチャの前にある２つのＢピクチャ（f_1 、f_2 ）によるカット点に対しては、Ｐピクチャを利用する結果の確認は行うことができない。f_3 のＩピクチャは参照を行わないからである。この部分で起こる誤検出の検出のためにもこのＩピクチャによる結果の確認が必要となる。

【0074】Ｉピクチャ（f_3 ）とひとつ前のＧＯＰにおけるＩピクチャ、それぞれのＤＣ画像に対する色ヒストグラム距離$D_{histarea}$（式（２））を調べ、次式（１１）を満す場合は間にあるＢピクチャから求めたカット点は誤検出であるとみなし、これを除去する。

【0075】

【数11】

$$D_{histarea} < threshold_I \qquad (11)$$

【0076】次にショットの代表フレームの選出について説明する。ここに代表フレームとはショットは動画全体に比べれば短い単位ではあるが、例えば５秒間のショットでは１５０枚（３０f_{ps}の場合）のフレームの集合であり、このままでは、比較、表示、特徴値の検出などの処理がしにくい。そこで、一般に、ショットを扱う際には、ショットの中からそのショットを代表するフレームを選び出し、この代表フレームによって比較、表示などの処理を行う。

【0077】この発明においては、ショットを統合しシーンを抽出する際に、ショット間の類似度を求めるために用いる。また、解析された動画構造をユーザに提示する際にショットの内容を簡単に示すためにも用いられる。したがって、ショットの内容を最もよく表しているフレームを選ぶことが必要となる。

【0078】ショットを扱っている従来の研究では、この代表フレームとして、機械的にショットの先頭のフレームあるいは中央のフレームを用いているものが多い。しかし、そのようにして選ばれたフレームはショットの内容をよく表すとは言い難い。そこでこの発明では、ショットに含まれるフレームの平均に最も近いフレームを代表フレームとして選ぶこととする。

【0079】また、ショットはほぼ動きがない場合だけではなく、

（１）ひとつのショットの途中でカメラの動き（パン・ズームなど）があるもの。

（２）カメラあるいは画像中のオブジェクトが動き続けているもの。

（３）動きが非常に激しいもの。

などの場合がある。このようなショットではひとつのフレームでショット全体を代表させるのは難しく、有用な情報を落とす危険がある。そこで、このようなショットは、複数の代表フレームによって表すこととする。

【0080】さらに、複数の代表フレームを選出するこ

とによって、カット点の検出ミスによる影響を少なくすることができる。つまり、カット点の検出ミスにより、本来複数であるショットがひとつにまとまってしまった場合、代表フレームをひとつだけ選出すると、本来、ただひとつのショットの情報だけしか使われないことになる。これに対し、内容に基づいて複数選出すれば、それぞれのショットの情報を捨てることなく、シーン抽出の際に生かせることになるわけである。

【００８１】必要最小限の代表フレームを選び出すため、まず、ショット中のフレームのクラスタリングを行う。この結果できた各クラスタからそれぞれもっとも平均に近いフレームを選び出し、これをショットの代表フレームとする。

【００８２】ただし、ショット内のすべてのフレームを代表フレームの候補とすると、ショットが長くなったときに処理量が増大する恐れがある。そこで、候補としてＩピクチャだけを用いる。これにより、選出のための処理量だけでなく、原動画像からの復号化のための処理量も削減することができる。また、ＭＰＥＧ１において一般に、符号化効率を上げるためにＩ、Ｐ、Ｂピクチャの量子化特性を変えることが多いため、Ｉピクチャが最も品質がよい場合が多いことも都合がよい。

【００８３】さらに、Ｉピクチャを完全に復号化するのではなく、前記で述べたＤＣ画像を用い、復号時の処理量削減を計る。

【００８４】具体的な処理手順は以下のようになる（図１３）。なお、Ｉピクチャがひとつも含まれないショットの場合、つまり非常に短いショットの場合はショット中で一番初めに現れるＰピクチャを、それも存在しない場合はＢピクチャを代表フレームとする。非常に短いショットの場合、ショット中での変化はほとんどないといえるから、このような機械的な処理で十分である。

【００８５】（１）ショットに含まれるＩピクチャを簡易復号化し、ＤＣ画像を取り出す。

【００８６】（２）ショット中、動きが少ない部分のＤＣ画像を初期クラスタとする。動きが少ないかどうかは、ＢおよびＰピクチャに含まれるＩＭＢの数を調べることによって行う（式（１２））。

【００８７】

【数１２】

$$N_I < threshold_{move} \qquad (12)$$

【００８８】（３）前記初期クラスタをもとにクラスタリングを行い、ショット中のＩピクチャをいくつかのクラスタへと分類する。クラスタリングは群平均法を使って行い、要素間の距離としては$D_{histarea}$（式（２））を用いる。クラスタリングは、クラスタ間の距離のうち最小のものが閾値を越えるまで行う。

【００８９】（４）クラスタリング終了後、各クラスタからそれぞれひとつずつ代表フレームを選び出す。まず、クラスタ内のＩピクチャのＤＣ画像を平均して、平均ＤＣ画像をつくる（式（１３））。

【００９０】

【数１３】

$$\overline{Y_{ij}} = \frac{1}{N}\sum_{n=1}^{N} Y_{ij}^{n}$$

$$\overline{Cb_{ij}} = \frac{1}{N}\sum_{n=1}^{N} Cb_{ij}^{n} \qquad (13)$$

$$\overline{Cr_{ij}} = \frac{1}{N}\sum_{n=1}^{N} Cr_{ij}^{n}$$

ここで，N はDC画像の数，$\overline{Y_{ij}}, \overline{Cb_{ij}}, \overline{Cr_{ij}}$ は平均DC画像の座標 (i,j) における輝度及び色差信号，$Y_{ij}^{n}, Cb_{ij}^{n}, Cr_{ij}^{n}$ は n 番目のDC画像 DC_n の座標 (i,j) における輝度及び色差信号を表す．

【００９１】（５）平均ＤＣ画像との距離D_{pixsum}（式（３））が一番小さいＤＣ画像DC_kをもつＩピクチャI_kを代表フレームとする。

【００９２】以上のようにしてショットの代表フレームが選出される。

【００９３】実際の処理では、ＩピクチャのＤＣ画像や、ＰおよびＢピクチャのマクロブロック情報を効率的に得るために、代表フレームの選出はカット点の検出と並行して行われる。

【００９４】たとえば会話のシーンなどでは、話者を交互に撮る場合が多いため、同じようなショットの繰り返しになる。このように、ひとつのシーンのなかには、似ているショットがいくつか含まれることが多い。この性質に着目してショットを統合し、シーンを抽出する。

【００９５】ショット間の類似度は、それぞれのショットの代表フレーム間の類似度ｓ（式（６））を用いる。ただし、ひとつのショットが複数の代表フレームを持つ場合もあるため、すべての代表フレームの組み合わせについての類似度を調べ、そのうちの最大値をショットの類似度とする（図１４）。

【００９６】ショット間の類似度からシーンを抽出する最も簡単な方法は、図１５のように、似ている（類似度が非情に高い）ショットが存在すれば、その間をすべて同じシーンとみなす方法である。

【００９７】しかし、このようにすれば、

（１）似ているか、似ていないかの閾値の設定が結果に大きく影響する

（２）類似度が非常に高いショットの組はないが、中程度の類似度のショットの組が多数ある、といった場合でも同じシーンとみなすことができず、柔軟性に欠けるといった問題点がある。

【００９８】そこで、ショット$shot_n$とショット$shot_{n+1}$が連続している（すなわち、同じシーンに属する）度合を表す結合度$connect_{n,n+1}$を式（１４）のように定義し、この結合度を用いてシーンの抽出を行う。

【００９９】

【数１４】

$$connect_{n,n+1} = 1 - \prod_{i=n-N+1}^{n} \prod_{j=n+1}^{i+N} (1 - s_{ij}) \qquad (14)$$

【０１００】ここで、Nは比較するショットの範囲を表す。s_{ij}は$shot_i$と$shot_j$の類似度である。

【０１０１】このように、結合度$connect_{n,n+1}$はショット$shot_n$とショット$shot_{n+1}$だけでなく、その付近のすべてのショット間の類似度s_{ij}から求められる。例えば、図１６において、$connect_{3,4}$はショット$shot_3$とショット$shot_4$の類似度s_{34}だけでなく、ショット$shot_2$とショット$shot_5$の類似度s_{25}も使って求められる。なぜなら、たとえショット$shot_3$とショット$shot_4$がまったく類似していなくても、ショット$shot_2$とショット$shot_5$が類似していれば、ショット$shot_3$とショット$shot_4$は同じシーンに属すると考えられるからである。

【０１０２】ただし、時間的に遠く離れているショット同士は、同じシーンに属する可能性が小さく、むしろ違うシーンに属するにもかかわらずたまたま類似度が高いショットが存在する可能性があり、このような原因による未検出をできるだけ防ぐため、比較するショットの範囲はNに制限する。

【０１０３】式（１３）によって得られる結合度$connect_{n,n+1}$の変化は、例えば図１７のようになる。

【０１０４】このような結合度の変化から、シーンチェンジを決定しシーンを抽出する。ここでは、変化のピークと谷の差が閾値$threshold_{SCENE}$より大きいとき、その谷となる結合度をもつカット点をシーンチェンジ点とする。

【０１０５】

【発明の効果】この発明によれば、与えられた動画像をハイブリッド符号化し、これを階層的構造へ分割し、分割されたショット間の類似度によりショットを統合してシーンを抽出するので、シーンの抽出が、迅速、正確に行われる効果がある。然してシーンからショット又はフレームを抽出するのは比較的容易であるから、結局動画像からシーン、ショット又はカットを短時間に、かつ正確に抽出し得る効果がある。

【０１０６】前記処理は総て現在使用されているハードに、適切なソフトを組み込むことにより自動化できるので、適切な入力指示により、所望のシーン、ショット又はカットを自動的に提供できる効果がある。

【０１０７】実験の結果によれば、表２の動画像を用いたカットの検出結果は表３の通りである。

【０１０８】

【表２】

表２　評価に用いた動画像

	題　　名	長さ(分)	フレーム数	画面サイズ	ソース
A	A Room with a View	31:52	45838	352×240	VideoCD
B	Kramer v.s. Kramer	30:09	43379	352×240	VideoCD
C	Stand by Me	30:32	43942	352×240	VideoCD

【０１０９】

【表３】

表３　カット点検出の結果

動画像	カット点の数	検出数	未検出数	誤検出数	検出率(%)	誤検出率(%)
A	272	270	2	11	99.2	4.04
B	267	253	14	0	94.7	0
C	377	259	18	0	95.2	0

【０１１０】またカット点検出の処理時間は表４の通り　　である。

【0111】
【表4】

表4　カット点検出の処理時間

動画像	処理時間(sec.)
A	110.44
B	107.09
C	98.74

【0112】更に単純なアルゴリズムによるカット点検出の処理時間は表5の通りである。
【0113】
【表5】

表5　単純なアルゴリズムによるカット点検出の処理時間

動画像	処理時間(sec.)
A	5160
B	5441
C	5161

【0114】次にカット点検出の結果得られたショットを用いて、シーン抽出を行った。結合度の変化を図18、19、20に示す。
【0115】図18、19、20の結合度からシーン抽出を行った結果が表6である。なお、シーンチェンジ点検出の閾値は、$threshold_{SCENE}$＝0．3とした。
【0116】表6から、いずれの動画像についてもシーンチェンジ点のうち75%以上を検出できており、高速性や本手法が意味解析や知識を使っていないことを考慮すると十分実用的であるといえる。なお、検出数の内1／4から1／3程度は、実際のシーンチェンジ点から1ショット分前または後ろにずれて検出されている。これは、シーンチェンジ点に隣接するショットに関して、その本来属するべきシーンのなかにそのショットへの類似度が高いショットが存在しない場合は結合度が低くなってしまうというアルゴリズム上の欠点による。
【0117】
【表6】

表6　シーンチェンジ点検出の結果

動画像	シーンチェンジ点の数	検出数	未検出数	誤検出数
A	17	17(4)	4	7
B	22	21(5)	1	11
C	24	18(6)	6	3

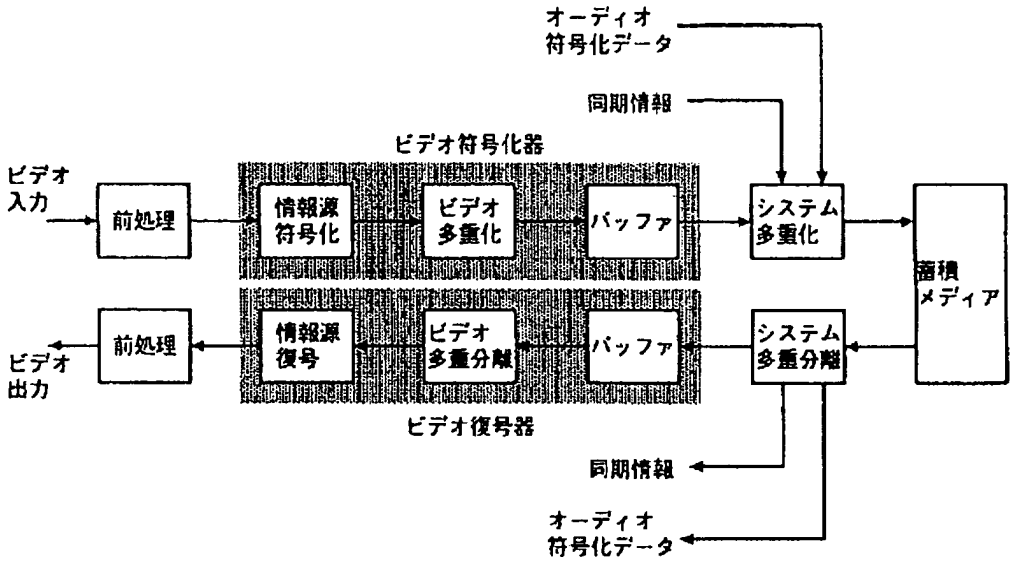
【0118】構造解析処理全体の処理時間は表7のようになっており、十分な高速性を保っていることがわかる。
【0119】
【表7】

表7　構造解析処理の処理時間

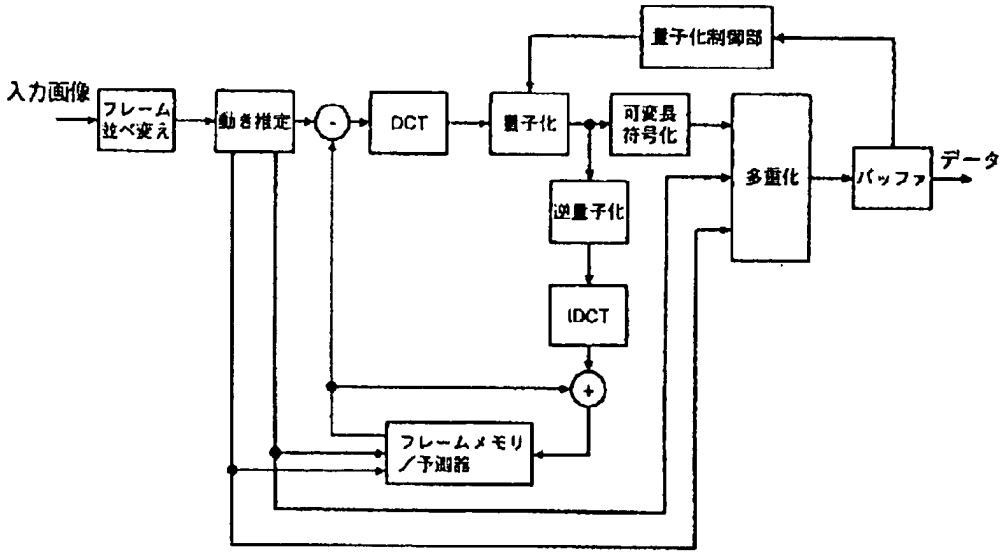
動画像	処理時間(sec.)
A	128.2
B	127.1
C	126.5

【図面の簡単な説明】
【図1】この発明のMPEG1符号化・復号システム。
【図2】同じくMPEG1ビデオ符号化器のブロック図。
【図3】同じくMPEG1ビデオ復号器の図。
【図4】同じくGOPの例示図。
【図5】同じく原画像およびストリーム上の画面の並びを示す図。
【図6】同じくMPEG1の階層構成図。
【図7】同じく構造解析処理の流れ図。
【図8】同じくDC画像の生成図。
【図9】同じくDC画像の例示図。
【図10】同じく類似度を求めるファジィ推論に用いるメンバシップ関数の形状図。
【図11】同じくGOPの例示図。
【図12】（a）同じく通常の参照図。
（b）同じく過去のPピクチャとの間にカット点がある場合を示す図。
（c）Bピクチャの間にカット点がある場合を示す図。
（d）同じく未来のPピクチャとの間にカット点がある場合の図。
【図13】同じく代表フレームの選出アルゴリズムを示す流れ図であって、
（a）DC画像を取り出す図。
（b）初期クラスタを決定する図。
（c）クラスタリングの図。
（d）クラスタ中のDC画像から平均画像を作る図。
（e）平均画像に最も近いものを代表フレームとする図。
【図14】同じくショット間の類似度を示す図。
【図15】同じく単純なシーン抽出の図。
【図16】同じくN＝3のときの結合度を示す図。
【図17】同じく結合度$connect_{n,\ n+1}$の変化の例示図。
【図18】同じく動画像Aの結合度の変化を示す例示図。
【図19】同じく動画像Bの結合度の変化を示す例示図。
【図20】同じく動画像Cの結合度の変化を示す図。

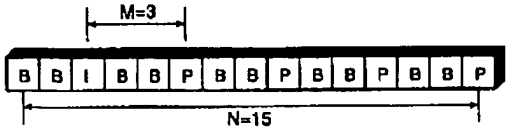
【図1】



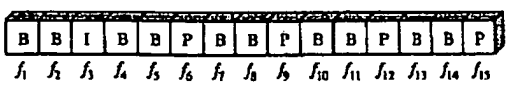
【図2】



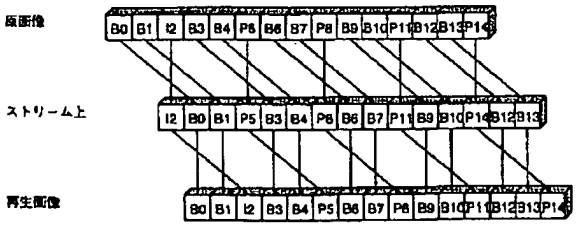
【図4】



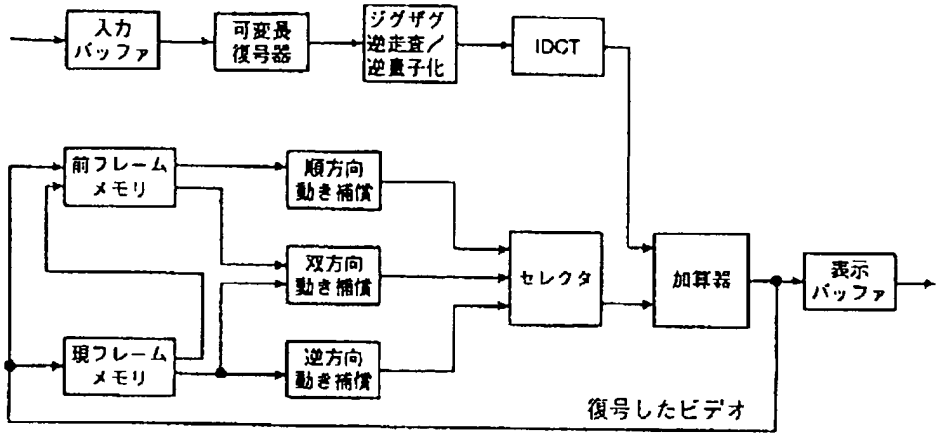
【図11】



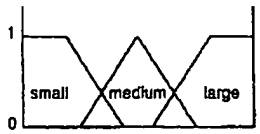
【図5】



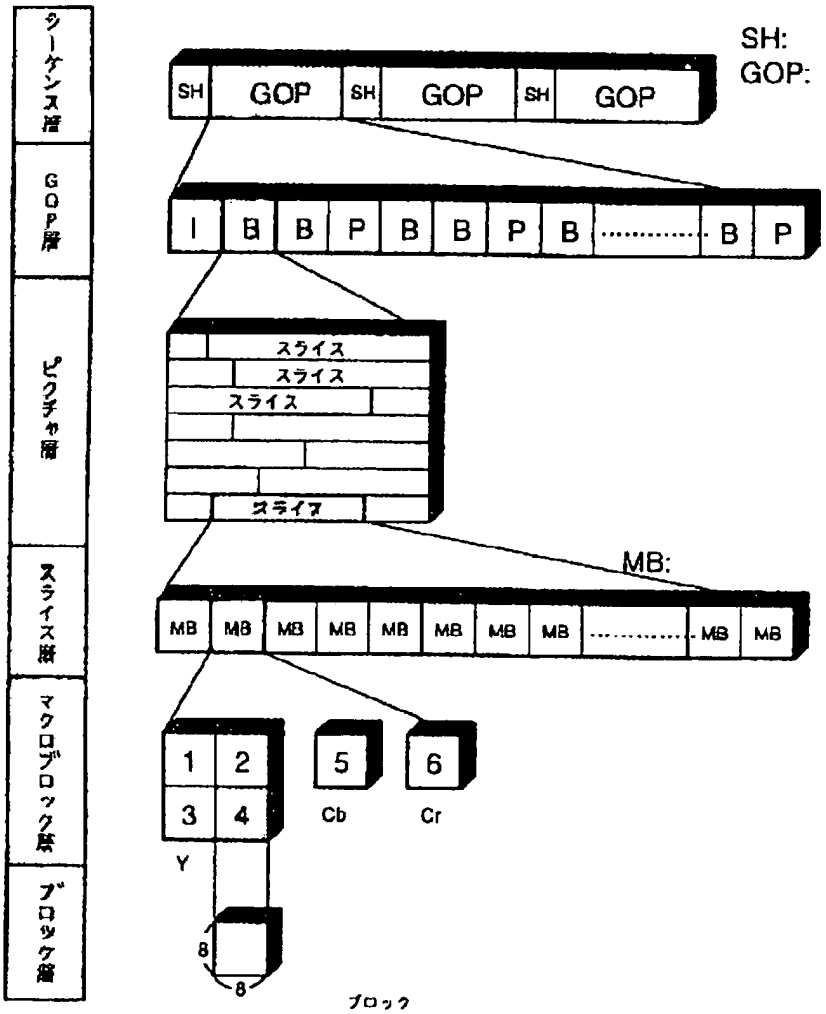
【図3】


入力バッファ
可変長復号器
ジグザグ逆走査／逆量子化
IDCT
前フレームメモリ
順方向動き補償
双方向動き補償
セレクタ
加算器
表示バッファ
現フレームメモリ
逆方向動き補償
復号したビデオ

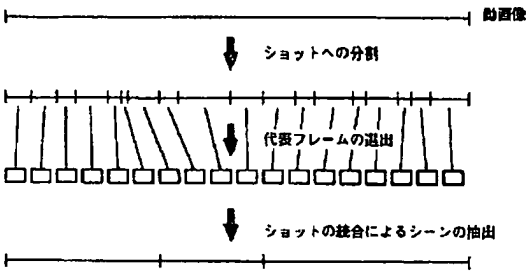
【図10】


1
0
small
medium
large

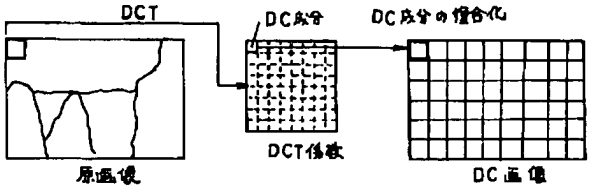
【図6】


シーケンス層
GOP層
ピクチャ層
スライス層
マクロブロック層
ブロック層
SH:
GOP:
SH
GOP
I
B
P
スライス
MB:
MB
1
2
3
4
5
6
Y
Cb
Cr
8
ブロック

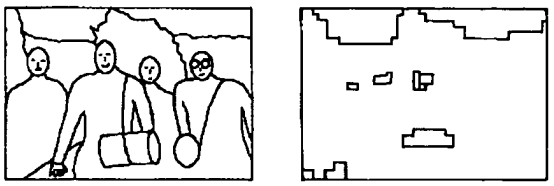
【図7】



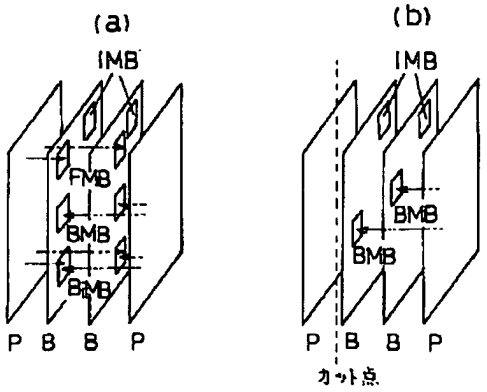
【図8】



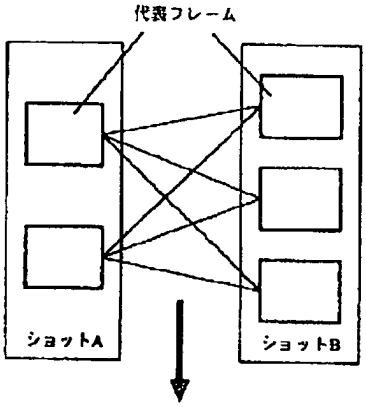
【図9】



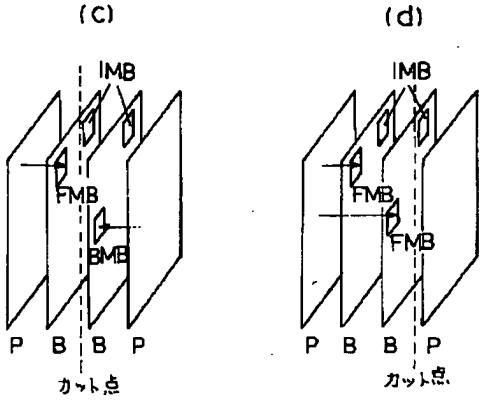
【図12】



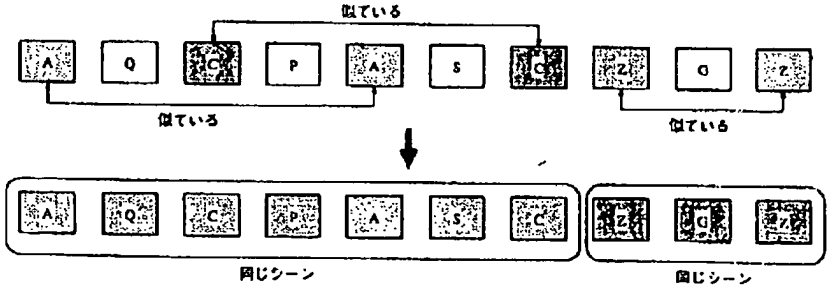
【図14】



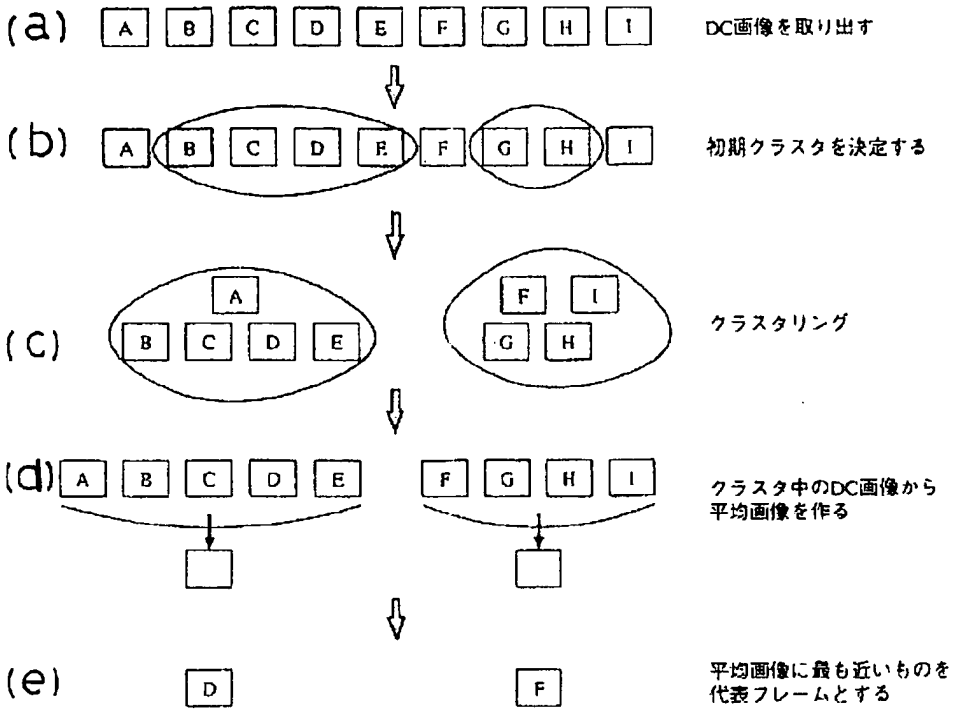
全ての組合せの代表フレーム間の類似度のうち
最大値をショット間の類似度とする



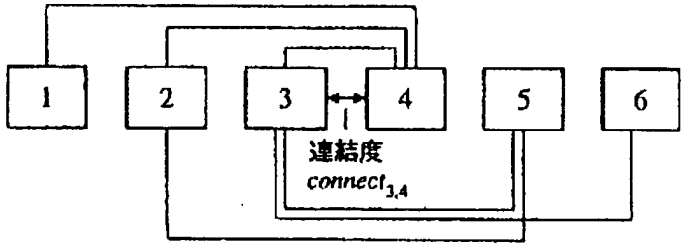
【図15】



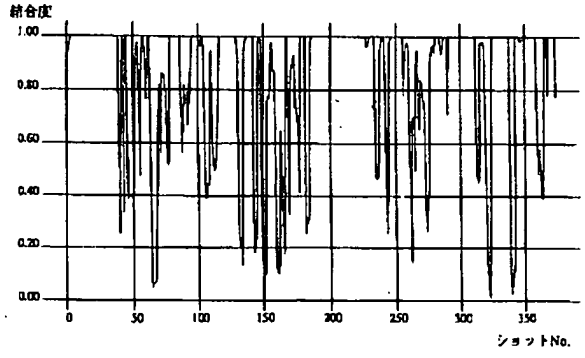
【図13】



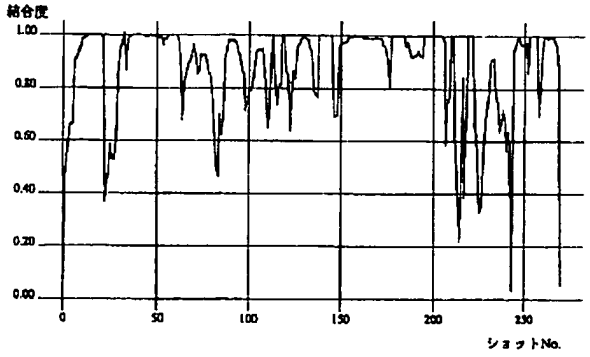
【図16】



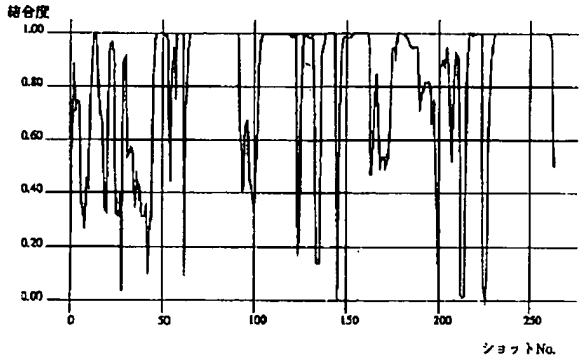
【図17】



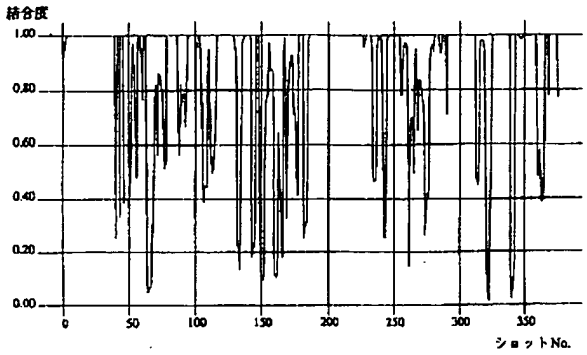
【図18】



【図19】


結合度
1.00
0.80
0.60
0.40
0.20
0.00
0
50
100
150
200
250
ショットNo.

【図20】


結合度
1.00
0.80
0.60
0.40
0.20
0.00
0
50
100
150
200
250
300
350
ショットNo.